

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of

Hideki KEMMOCHI

Serial No.: NEW APPLICATION

Group Art Unit:

Filed: March 3, 2004

Examiner:

For: SINGING VOICE SYNTHESIZING APPARATUS WITH SELECTIVE USE OF  
TEMPLATES FOR ATTACK AND NON-ATTACK NOTES

**CLAIM FOR PRIORITY**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

Sir:

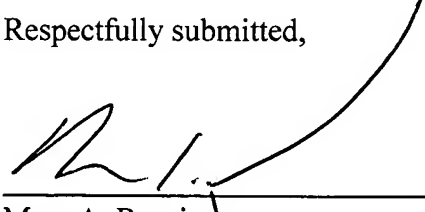
The benefit of the filing date of the following prior foreign application filed in the following country is hereby requested for the above-identified application and the priority provided in 35 U.S.C. § 119 is hereby claimed:

JAPAN 2003-055898 March 3, 2003

In support of this claim, a certified copy of said original foreign application is filed herewith. It is requested that the file of this application be marked to indicate that the requirements of 35 U.S.C. 119 have been fulfilled and that the Patent and Trademark Office kindly acknowledge receipt of this document.

Respectfully submitted,

03/03/04  
Date

  
\_\_\_\_\_  
Marc A. Rossi  
Registration No. 31,923

Attorney Docket: YAMA:067

ROSSI & ASSOCIATES  
P.O. Box 826  
Ashburn, VA 20146-0826

日本国特許庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2003年 3月 3日  
Date of Application:

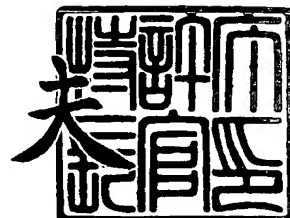
出願番号 特願2003-055898  
Application Number:  
[ST. 10/C]: [JP 2003-055898]

出願人 ヤマハ株式会社  
Applicant(s):

2003年 9月25日

特許庁長官  
Commissioner,  
Japan Patent Office

今井 康



【書類名】 特許願

【整理番号】 C31121

【提出日】 平成15年 3月 3日

【あて先】 特許庁長官 殿

【国際特許分類】 G10H 1/00

【発明の名称】 歌唱合成装置、歌唱合成プログラム

【請求項の数】 6

【発明者】

【住所又は居所】 静岡県浜松市中沢町 1 0 番 1 号 ヤマハ株式会社内

【氏名】 剣持 秀紀

【特許出願人】

【識別番号】 000004075

【氏名又は名称】 ヤマハ株式会社

【代理人】

【識別番号】 100098084

【弁理士】

【氏名又は名称】 川▲崎▼ 研二

【手数料の表示】

【予納台帳番号】 038265

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 歌唱合成装置、歌唱合成プログラム

【特許請求の範囲】

【請求項 1】 立ち上がり音が各種の表情を持つために必要な特徴量の時間的変化を表情毎に定義したテンプレートデータを記憶すると共に、非立ち上がり音が各種の表情を持つために必要な特徴量の時間的変化を表情毎に定義したテンプレートデータを記憶する記憶手段と、

歌詞を構成する一連の音韻と、各音韻に対する表情とを指定する音声情報を入力する入力手段と、

前記音声情報により指定された一連の音韻のうち立ち上がり音については、前記記憶手段に記憶された立ち上がり音用のテンプレートデータのうち、前記音声情報により指定された当該立ち上がり音の表情に対応したテンプレートデータを用いて音声を合成し、前記一連の音韻のうちの非立ち上がり音については、前記記憶手段に記憶された非立ち上がり音用のテンプレートデータのうち、前記指定された当該非立ち上がり音の表情に対応したテンプレートデータを用いて音声を合成する合成手段と、

を備える歌唱合成装置。

【請求項 2】 請求項 1 記載の歌唱合成装置において、

前記入力手段に入力される音声情報は、

歌詞を構成する一連の音韻と、当該各音韻の発音タイミングとを関連付けた音韻トラックを含み、

前記合成手段は、

前記音声情報により指定された各音韻が、時系列的に先行して発音される先行音韻を有し且つ当該先行音韻との発音タイミングの差が所定期間未満である音韻であるかを前記音韻トラックに基づいて夫々判断し、前記先行音韻を有し且つ当該先行音韻との発音タイミングの差が所定期間未満である音韻を非立ち上がり音と特定すると共に、前記先行音韻を有さず又は前記先行音韻との発音タイミングの差が所定期間以上である音韻を立ち上がり音と特定する、

歌唱合成装置。

【請求項 3】 請求項 1 記載の歌唱合成装置において、  
前記入力手段に入力される音声情報は、  
各音韻に付する音高の遷移を前記各音韻の発音タイミングに対応させて記録したピッチトラックを含み、  
前記合成手段は、  
前記音声情報により指定された各音韻の発音タイミングから所定期間遡ったタイミングまでの先行区間に音高を指定する値が記録されているかを前記ピッチトラックに基づいて夫々判断し、前記先行区間に音高を指定する値が記録されている音韻を非立ち上がり音と特定すると共に、前記先行区間に音高を指定する値が記録されていない音韻を立ち上がり音と特定する、  
歌唱合成装置。

【請求項 4】 立ち上がり音が各種の表情を持つために必要な特徴量の時間的変化を表情毎に定義したテンプレートデータを記憶すると共に、非立ち上がり音が各種の表情を持つために必要な特徴量の時間的変化を表情毎に定義したテンプレートデータを記憶したコンピュータを動作させる歌唱合成プログラムであって、

歌詞を構成する一連の音韻と、各音韻に対する表情とを指定する音声情報を入力する入力工程と、

前記音声情報により指定された一連の音韻のうち立ち上がり音については、前記記憶された立ち上がり音用のテンプレートデータのうち、前記音声情報により指定された当該立ち上がり音の表情に対応したテンプレートデータを用いて音声を合成し、前記一連の音韻のうちの非立ち上がり音については、前記記憶された非立ち上がり音用のテンプレートデータのうち、前記指定された当該非立ち上がり音の表情に対応したテンプレートデータを用いて音声を合成する合成工程と、  
をコンピュータに実行させる歌唱合成プログラム。

【請求項 5】 請求項 4 記載の歌唱合成プログラムにおいて、  
前記入力工程で入力される音声情報は、  
歌詞を構成する一連の音韻と、当該各音韻の発音タイミングとを関連付けた音韻トラックを含み、

前記合成工程は、

前記音声情報により指定された各音韻が、時系列的に先行して発音される先行音韻を有し且つ当該先行音韻との発音タイミングの差が所定期間未満である音韻であるかを前記音韻トラックに基づいて夫々判断し、前記先行音韻を有し且つ当該先行音韻との発音タイミングの差が所定期間未満である音韻を非立ち上がり音と特定すると共に、前記先行音韻を有さず又は前記先行音韻との発音タイミングの差が所定期間以上である音韻を立ち上がり音と特定する工程を含む、

歌唱合成プログラム。

【請求項 6】 請求項 4 記載の歌唱合成プログラムにおいて、

前記入力工程で入力される音声情報は、

各音韻に付する音高の遷移を前記各音韻の発音タイミングに対応させて記録したピッチトラックを含み、

前記合成工程は、

前記音声情報により指定された各音韻の発音タイミングから所定期間遡ったタイミングまでの先行区間に音高を指定する値が記録されているかを前記ピッチトラックに基づいて夫々判断し、前記先行区間に音高を指定する値が記録されている音韻を非立ち上がり音と特定すると共に、前記先行区間に音高を指定する値が記録されていない音韻を立ち上がり音と特定する工程を含む、

歌唱合成プログラム。

【発明の詳細な説明】

【 0 0 0 1 】

【発明の属する技術分野】

本発明は、歌唱合成装置に係り、特に、表情を付けた自然な歌唱音を合成する歌唱合成装置に関する。

【 0 0 0 2 】

【従来の技術】

従来より、実際に発声した音声进行分析して得た各種パラメータを準備し、これらパラメータを組み合わせて歌唱音声を合成するといったことが行われている。このような技術の一つを開示した文献に、特開 2 0 0 2 - 2 6 8 6 5 9 がある。

(特許文献 1 参照)。

#### 【0003】

同文献に開示する技術は以下のようなものである。まず、各音韻のフォルマントなどを特徴付けるパラメータを格納したデータベースと、このパラメータに時系列的な変化を与えるためのテンプレートデータを格納したデータベースとを予め準備する。

#### 【0004】

また、歌詞の音韻を時系列的に指定する音韻トラック、歌い出し部分や音符の遷移部分等とその時刻を指定するノートトラック、各音韻の音高を指定するピッチトラック、各時刻における音声の強さを指定するダイナミクストラック、及び各時刻における唇開度を指定するオープニングトラックを備えるスコアデータを予め準備する。

#### 【0005】

演奏を行う際は、スコアデータの各トラックからパラメータを読み出し、このパラメータに上記テンプレートデータを適用することによって時間毎の微小な変動を持たせた最終的なパラメータを取得し、この最終的なパラメータから音声合成を行う。

#### 【0006】

音声合成を行うために準備するパラメータやテンプレートデータの種類の多様であり、より細かな種類のパラメータやテンプレートデータを用意してこれらから音声を合成することによって、多様で且つ人間の発声に近い歌唱音声を合成することができる。

#### 【0007】

##### 【特許文献 1】

特開 2002-268659 号公報

#### 【0008】

##### 【発明が解決しようとする課題】

上述したような、多様で且つ人間の発声に近い歌唱音声を合成するために用意することが望ましいテンプレートデータの一つとして、アクセント、ポルタメン

トといった表情に関するものがある。表情を付けて歌唱する場合とそうでない場合とでは、音韻のフォルマントやピッチの変動する態様は変わってくるし、表情の種別によってもこの変動の態様は変わってくるからである。したがって、異なる各表情に対応するテンプレートを用意し、ユーザに指定された表情のテンプレートをこの表情が指定された区間に適用するようにすれば、より一層多様な歌唱音声の合成が実現できる。

#### 【0009】

しかし、このような表情を達した音声合成を実現するためには克服すべき問題もあった。同種の表情を付けて歌唱するにしても、その表情を付ける区間の前につながった音符がある場合とそうでない場合とでは、音韻のフォルマントやピッチの変動する態様が異なるため、表情を付ける区間の前につながった音符がある場合とそうでない場合とでは異なるテンプレートデータを適用しなければ適切な歌い回しを再現できない。

#### 【0010】

表情を付する区間の前につながった音符があるかどうかにより2つの異なるテンプレートデータを用意すること自体は、実際にそのような条件で歌唱した音声それぞれ分析すれば比較的簡単に実現できる。しかし、歌唱データを作成するユーザは、ある区間に表情を付するために2つの異なるテンプレートデータのうちのいずれのテンプレートデータを適用するかを逐一判断するという煩わしい作業を強いられることになる。

#### 【0011】

この発明は、このような問題を克服するために案出されたものであり、歌唱データを作成するユーザが特定区間に表情を与えようとする場合に、その区間の前に連続する音符があるかといったことを意識しなくても、適切な表情テンプレートが適用される歌唱合成装置を提供することを目的とする。

#### 【0012】

##### 【課題を解決するための手段】

本発明は、立ち上がり音が各種の表情を持つために必要な特徴量の時間的变化を表情毎に定義したテンプレートデータを記憶すると共に、非立ち上がり音が各



種の表情を持つために必要な特徴量の時間的変化を表情毎に定義したテンプレートデータを記憶する記憶手段と、歌詞を構成する一連の音韻と、各音韻に対する表情とを指定する音声情報を入力する入力手段と、上記音声情報により指定された一連の音韻のうち立ち上がり音については、上記記憶手段に記憶された立ち上がり音用のテンプレートデータのうち、上記音声情報により指定された当該立ち上がり音の表情に対応したテンプレートデータを用いて音声を合成し、上記一連の音韻のうちの非立ち上がり音については、上記記憶手段に記憶された非立ち上がり音用のテンプレートデータのうち、上記指定された当該非立ち上がり音の表情に対応したテンプレートデータを用いて音声を合成する合成手段とを備える、歌唱合成装置を提供する。

#### 【0013】

ここで、「立ち上がり音」とは、何も歌唱していない状態から歌唱を開始するという起点に位置する音韻をいう。

#### 【0014】

また、本発明は、上記入力手段に入力される音声情報に、歌詞を構成する一連の音韻と、当該各音韻の発音タイミングとを関連付けた音韻トラックを含み、上記合成手段が、上記音声情報により指定された各音韻が、時系列的に先行して発音される先行音韻を有し且つ当該先行音韻との発音タイミングの差が所定期間未満である音韻であるかを上記音韻トラックに基づいて夫々判断し、上記先行音韻を有し且つ当該先行音韻との発音タイミングの差が所定期間未満である音韻を非立ち上がり音と特定すると共に、上記先行音韻を有さず又は上記先行音韻との発音タイミングの差が所定期間以上である音韻を立ち上がり音と特定する歌唱合成装置を提供する。

#### 【0015】

また、本発明は、上記入力手段に入力される音声情報に、各音韻に付する音高の遷移を上記各音韻の発音タイミングに対応させて記録したピッチトラックを含み、上記合成手段が、上記音声情報により指定された各音韻の発音タイミングから所定期間遡ったタイミングまでの先行区間に音高を指定する値が記録されているかを上記ピッチトラックに基づいて夫々判断し、上記先行区間に音高を指定す

る値が記録されている音韻を非立ち上がり音と特定すると共に、上記先行区間に音高を指定する値が記録されていない音韻を立ち上がり音と特定する歌唱合成装置を提供する。

【0016】

【発明の実施の形態】

#### A 第1の実施の形態

以下、図面を参照しつつ、本発明の第1の実施の形態について説明する。本実施形態の特徴は、表情を付けて歌唱することを指示された区間の最初の音韻のコンテキストを判別し、このコンテキストの種類に応じた適切な表情テンプレートデータを当該区間に適用する点にある。

【0017】

テンプレートデータとは、歌唱音を特徴付ける値を如何なる態様で時間的に変化させるかを定義したデータを意味するが、その具体的内容は後に詳述する。また、コンテキストとは、発音する音韻の前後関係を意味する。そして、本実施形態におけるコンテキストには、ノートアタックとノートトランジッションの2つの種類がある。ノートアタックとは、何も歌唱していない状態から歌唱を開始するという起点に位置することをいい、ノートトランジッションとは、ノートアタックでないこと、即ち、発音中の音韻がある音韻から後発の音韻へと遷移する際の後発に位置することをいう。

【0018】

ある特定の区間を特定の表情を付して歌唱した場合、同じ種類の表情を付するにしても、その特定の区間の最初の音韻がノートアタックかノートトランジッションかによって歌い回しは異なる。このような歌いまわしの違いを適切に再現するべく、本実施形態では、後に詳述するような動作によって、表情を付ける区間の最初の音韻のコンテキストに応じた適切なテンプレートデータを自動的に選出し適用することになっている。

【0019】

ここで、本実施形態における「音韻」について定義しておく。本実施形態においては、音高を付けて発音することができる音素又は音素列をひとつの「音韻」

の単位として説明を行う。具体的には、子音の音素とこれに引き続く母音の音素とをつなぎ合わせた音素列（例えば「か」）、或いは母音のみの音素（例えば「あ」）を一つの「音韻」と考える。

#### 【0020】

##### <実施形態の構成>

図1は、本実施形態に係る歌唱合成装置の物理的構成を示すブロック図である。同図に示すように、当該歌唱合成装置は、CPU100、ROM110、RAM120、タイマ130、ディスプレイ140、マウス150、キーボード160、DAC（D/Aコンバータ）170、サウンドシステム180、MIDIインターフェース190、記憶装置200およびバスを備えている。なお、ディスプレイ140、マウス150、キーボード160および記憶装置200の各種インターフェースは図示を省略してある。

#### 【0021】

記憶装置200は、例えば、ハードディスク装置（HDD）であり、OS（オペレーティングシステム）および各種のアプリケーションプログラムを記憶している。なお、記憶装置200は、CD-ROM装置、光磁気ディスク（MO）装置、デジタル多目的ディスク（DVD）装置等であってもよい。CPU100は、例えば記憶装置200にインストールされたOS（オペレーティングシステム）を実行し、ディスプレイ140の表示とマウス150の操作による所謂GUI（グラフィカル・ユーザズ・インターフェース）をユーザに提供する。また、CPU100は、GUIを介してユーザからアプリケーションプログラムの実行指示を受け取り、指示されたアプリケーションプログラムを記憶装置200から読み出して実行する。記憶装置200に記憶されたアプリケーションプログラムとして、歌唱合成プログラムがある。この歌唱合成プログラムは、本実施形態に特有の動作をCPU100に実行させる。RAM120は、その際にワーキングエリアとして使用される。

#### 【0022】

MIDIインターフェース190は、他のMIDI機器から歌唱データを受け取り、あるいは歌唱データをMIDI機器に出力する機能を有する。

**【0023】**

図2は、本実施形態に係る歌唱合成装置の論理的構成を示すブロック図である。同図左側には、CPU100がその役割を担う各部の構成を示し、同図右側には、記憶装置200に編成されたデータベースの構成を示す。

**【0024】**

まず、記憶装置200にインストールされた歌唱合成プログラムを実行することによって、CPU100は、インターフェース制御部101、スコアデータ生成部102、コンテキスト判別部104、スコアデータ更新部103、特徴パラメータ発生部105、及び歌唱合成部106としての役割を担う。

**【0025】**

インターフェース制御部101は、ディスプレイ140に表示する歌唱データ編集画面の制御を司る。ユーザは、この歌唱データ編集画面を参照することにより歌唱スコアデータを編集するためのデータ入力を行う。この歌唱スコアデータは、時間とともに変化する歌唱音のフレーズを複数のトラックによって表した歌唱データである。なお、この歌唱データ編集画面の構成及び歌唱スコアデータは後に詳述する。

**【0026】**

スコアデータ生成部102は、ユーザからの入力データを用いて歌唱スコアデータを生成する。コンテキスト判別部104は、上記歌唱スコアデータによって表された各音韻のコンテキストを判別する。スコアデータ更新部103は、上記コンテキスト判別部104による判別結果に基づいて、上記歌唱スコアデータにコンテキストデータを追加する。このコンテキストデータは、当該歌唱スコアデータによって表されている各音韻がノートアタックの音かそれともノートランジションの音かを特定するためのデータである。

**【0027】**

特徴パラメータ発生部105は、歌唱スコアデータおよびコンテキストデータに基づいて、発生すべき歌唱音の特徴パラメータを発生して、歌唱合成部106に出力する。この特徴パラメータは、例えば、励起波形スペクトルのエンベロープ、励起レゾナンス、フォルマント、差分スペクトルの4つに分類することが出

来る。これらの4つの特徴パラメータは、実際の人間の音声等（オリジナルの音声）を分析して得られる調和成分のスペクトル・エンベロープ（オリジナルのスペクトル）を分解することにより得られるものである。

#### 【0028】

歌唱合成部106は、歌唱スコアデータの各トラックに記録された値と、上記特徴パラメータとによってデジタル音楽信号を合成する。

#### 【0029】

次に、図2右側に示す各種データベースを説明する。Timbreデータベース210は、音韻名、ピッチを異にする各特徴パラメータを集めたデータベースである。ある時刻における音声は特徴パラメータ（励起スペクトル、励起レゾナンス、フォルマント、差分スペクトルのセット）で表現でき、同じ音声でもピッチが異なればこれを表現する特徴パラメータも異なる。このTimbreデータベース210は、インデックスとして音韻名、ピッチを持つ。従って、CPU100は、上記歌唱スコアデータの音韻トラック及びピッチトラックに属するデータをキーとして、ある時刻 $t$ における特徴パラメータを読み出すことができる。

#### 【0030】

表情テンプレートデータベース200は、音韻に対して表情付けを行うのに使用するテンプレートデータを集めたデータベースである。本実施形態において、音韻に付ける表情には、アクセント、ソフト、レガート、ポルタメントがある。本実施形態では、これらの表情を音韻に付けるため、音韻に対応した音声波形の特徴パラメータやピッチを時間経過に伴って変化させる。テンプレートデータは、歌唱音を特徴付けるパラメータを如何なる態様で時間的に変化させるかを定義したデータであることは上述したところであるが、ここにいう「歌唱音を特徴付けるパラメータ」とは、具体的には、特徴パラメータ $P$ とピッチ $Pitch$ である。本実施形態におけるテンプレートデータは、時間 $t$ の関数として表された特徴パラメータ $P$ とピッチ $Pitch$ とを一定時間 $\Delta t$ 間隔でサンプリングしたデジタル値のシーケンスと、特徴パラメータ $P$ とピッチ $Pitch$ の区間長 $T$  (sec.) の組により構成されるものであり、以下の式(A)により表すことができる。

## 【数 1】

Template=[P (t), Pitch(t), T]… (A)

## 【0031】

なお、 $t=0$ 、 $\Delta t$ 、 $2\Delta t$ 、 $3\Delta t$ 、… $T$ であり、本実施形態では、 $\Delta t$ は 5ms とする。 $\Delta t$ を小さくすると時間分解能がよくなるので音質は良くなるがデータベースのサイズが大きくなり、逆に $\Delta t$ を大きくすると音質が悪くなるがデータベースのサイズは小さくなる。 $\Delta t$ を決定する際には音質とデータベースのサイズとの優先度を考慮して決定すればよい。

## 【0032】

表情テンプレートデータベース200は、ノートアタック表情テンプレートデータベース220とノートランジッション表情テンプレートデータベース230とに分かれている。

## 【0033】

ここで、ノートアタック表情テンプレートデータベース220は、ノートアタックの音から始まる区間の表情付けに使用するテンプレートデータを集めたデータベースである。このノートアタック表情テンプレートデータベース220は、表情付けの種類により、アクセントテンプレートデータベース221とソフトテンプレートデータベース222とに区分されている。そして、ノートアタック表情テンプレートデータベース220内の各テンプレートデータベースには、予め想定した複数の音韻および複数の代表ピッチの全組み合わせについて、図3に示すように、音韻名と代表ピッチをインデックスとするテンプレートデータが用意されている。ここで、図2に示すように、ノートアタック表情テンプレートデータベース220には、レガート及びポルタメントを指定した区間に適用されるテンプレートデータのデータベースが用意されていないが、これは、音の立ち上がり時にレガート又はポルタメントをかけて発声するということがありえないからである。

## 【0034】

一方、ノートランジッション表情テンプレートデータベース230は、ノートランジッションの音から始まる区間の表情付けに使用する表情テンプレートデ

ータを集めたデータベースである。このノートトランジション表情テンプレートデータベース230は、表情付けの種類により、アクセントテンプレートデータベース231とソフトテンプレートデータベース232とレガートテンプレートデータベース233とポルタメントテンプレートデータベース234とに区分されている。ノートトランジション表情テンプレートデータベース230内の各テンプレートデータベースには、予め想定された複数の先発音韻名、複数の後発音韻名、複数の代表ピッチの全組み合わせについて、図4に示すように、先発音韻名、後発音韻名、代表ピッチをインデックスとするテンプレートデータが用意されている。

#### 【0035】

表情テンプレートデータベース200を構成するテンプレートデータは、後に詳述する歌唱データ編集画面において、アクセント、ソフト（やわらかく）、レガート（なめらかに）、ポルタメントといった表情の指定が行なわれた区間に適用される。

#### 【0036】

音韻テンプレートデータベース240は、音韻テンプレートデータを格納している。この音韻テンプレートデータは、上記歌唱スコアデータにおける音韻と音韻との遷移区間に適用するデータである。人間が2つの音韻を連続して発する場合には、突然変化するのではなくゆるやかに移行していく。例えば「あ」という母音の後に区切りを置かないで連続して「え」という母音を発音する場合には、最初に「あ」が発音され、「あ」と「え」の中間に位置する発音を経て「え」に変化する。従って、音韻の結合部分が自然になるように歌唱合成を行うには、ある言語において組合せ可能な音韻の組合せについて、結合部分の音声情報を何らかの形で持つことが好ましい。これを考慮し、本実施形態では、音韻が遷移する区間における、特徴パラメータとピッチの変動量をテンプレートデータとして準備し、歌唱スコアデータにおける音韻の遷移区間にこのテンプレートデータを適用することによって、より実際の歌唱に近い音声の合成を実現している。

#### 【0037】

この音韻テンプレートデータは、上記表情テンプレートデータと同様に、特徴

パラメータ  $P$  とピッチ  $Pitch$  のペアが一定時間ごとに並んだシーケンス、及びその区間の長さ  $T$  (sec.) の組であり、上記算出式 (A) で表すことが出来る。但し、上記表情テンプレートデータは時系列にしたがって変動する特徴パラメータ及びピッチの絶対値そのものを持つ構造となっているのに対し、この音韻テンプレートデータは、時間毎の特徴パラメータ及びピッチの変動量を持つ構造になっている。これは、後に詳述するように、表情テンプレートデータと音韻テンプレートデータとでは適用の仕方が異なるからである。

#### 【0038】

状態テンプレートデータベース 250 は、状態テンプレートデータを格納する。状態テンプレートデータは、上記歌唱スコアデータにおける、音韻の立ち上がり部分、及び音韻の遷移部分に全般的に適用するデータである。ある音韻を一定のピッチで発声しようとしたときの立ち上がり部分を解析すると、振幅が徐々に大きくなり一定のレベルで安定していくことが明らかになっている。また、2つの音符を連続して間を置かずに歌唱するときは、ピッチ及び特徴パラメータが微小な揺らぎを持ちながら変動していくことが明らかになっている。これを考慮し、本実施形態では、音韻の立ち上がりや遷移区間における特徴パラメータとピッチの変動量をテンプレートデータとして準備し、歌唱スコアデータにおける音韻の立ち上がりや遷移区間にこのテンプレートデータを適用することによって、より実際の歌唱に近い音声の合成を実現している。

#### 【0039】

この状態テンプレートデータも、特徴パラメータ  $P$  とピッチ  $Pitch$  のペアが一定時間ごとに並んだシーケンス、及びその区間の長さ  $T$  (sec.) の組であり、上記算出式 (A) で表すことが出来る。そして、上記音韻テンプレートデータと同様に、時間毎の特徴パラメータ及びピッチの変動量を持つ構造になっている。

#### 【0040】

##### <実施形態の動作>

以下、上記構成を有する歌唱合成装置の動作について説明する。図5は、当該歌唱合成装置の動作の概要を示すフローチャートである。



**【0041】**

CPU100は、GUIを介して歌唱合成の実行指示を受け取ると、記憶装置200から歌唱合成プログラムを読み出して実行する。この歌唱合成プログラムの実行過程では、図5に示す処理が行われる。まず、歌唱合成プログラムを構成するモジュールの1つであるインターフェース制御部101が、ディスプレイ140に歌唱データ編集画面を表示する(S110)。図6に歌唱データ編集画面を示す。歌唱データ編集画面のウィンドウ600は、ノートデータをピアノロール形式で表示するイベント表示領域601を備えている。イベント表示領域601の右側には、イベント表示領域601の表示画面を上下にスクロールするためのスクロールバー606が設けられている。イベント表示領域601の下側には、イベント表示領域601の表示画面を左右にスクロールするためのスクロールバー607が設けられている。

**【0042】**

イベント表示領域601の左側にはピアノの鍵盤を模した鍵盤表示602（ピッチを示す座標軸）が表示され、イベント表示領域601の上側には楽曲の先頭からの小節位置を示す小節表示604が表示される。603はピアノロール表示領域であり、鍵盤表示602で示されるピッチの小節表示604で示される時間位置にノートデータを横長の矩形（バー）で表示している。バーの左端位置は発声開始タイミングを示し、バーの長さは発声継続時間を示し、バーの右端位置は発声終了タイミングを示している。

**【0043】**

ユーザは、所望のピッチ及び時間位置に対応した表示画面上の位置にマウスポインタを移動してクリックし、発声開始位置を特定する。そして、ドラッグ操作により発声開始位置から発声終了位置に至るノートデータのバー（以下ノートバーという）をイベント表示領域601に形成し、その後、マウス150をドロップする。例えば、ノートバー611を形成するためには、第53小節目の第1拍め先頭の位置にマウスポインタを位置決めしてマウス150をクリックし、1拍後までドラッグすればよい。

**【0044】**

上記ドラッグアンドドロップによってノートバーを形成したユーザは、当該ノートバーに割り当てる歌詞の入力操作と、任意の指定である表情の入力操作とを行う。

#### 【0045】

歌詞の入力操作は、上記形成したノートバーにマウスポインタを移動しマウス 150 を右クリックすることによって、図 7 の拡大図に示すような歌詞入力欄をノートバーの上部に表示させ、キーボード 160 によって当該入力欄へ歌詞を入力する。

#### 【0046】

一方、表情の入力操作は、上記形成したノートバーにマウスポインタを移動しマウス 150 を左クリックすることによって、図 8 に示すような表情選択リストをノートバーの下部にプルダウン表示させ、そのリストの中から当該ノートバーに割り当てる表情を選択する。表情選択リストに表示される表情は、アクセント、やわらかく（ソフト）、レガート、ポルタメントの 4 つである。

#### 【0047】

ここで、同じピッチで複数の音韻を間を置かずに歌唱する場合でも、ユーザは、図 9 の拡大図に示すように同じピッチのノートバーを複数に分けて形成しなければならない。さもなければ、前の音韻をどこまで伸ばして、後の音韻をどこから発声するのか特定することができないからである。また、異なるピッチで単一の音韻を歌唱させるとき、ユーザは、図 10 の拡大図に示すように、まず、異なるピッチのノートバーを別々に形成し、先の音韻の歌詞を入力し、後の音韻の歌詞として「-」（ハイフン）を入力しなければならない。

#### 【0048】

このような操作によって、楽曲を演奏するのに必要なノートバー、歌詞、及び表情を入力したユーザは、図示しない歌唱音声出力ボタンをクリックする。

#### 【0049】

この歌唱音声出力ボタンをクリックされると、スコアデータ生成部 102 が、この入力されたノートデータ、表情に基づいて歌唱スコアデータを生成する（S120）。

## 【0050】

図11は、スコアデータ生成部102が生成する歌唱スコアデータの一例を示す概念図である。この歌唱スコアデータは、音韻トラック、ピッチトラック、表情トラックによって構成されている。

## 【0051】

音韻トラックには、音韻名と、その発音継続時間が記録される。この音韻トラックには、上記歌唱データ編集画面で各ノートバーに割り当てた歌詞が反映される。

## 【0052】

ピッチトラックには、各時刻において発音すべき音声の基本周波数が記録される。このピッチトラックには、上記歌唱データ編集画面の各ノートバーの縦座標が反映される。なお、実際に発音される音声のピッチはこのピッチトラックに記録されているピッチ情報に他の情報を作用させて算出されるので、実際に発音されているピッチと、ここに記録されているピッチとは異なる場合がある。

## 【0053】

表情トラックには、特定の音韻について指定された表情と、その表情の継続時間が記録される。表情には、「アクセント」を示す「A」と、「やわらかく（ソフト）」を示す「S」と、「なめらかに（レガート）」を示す「R」と、「ポルタメント」を示す「P」とがある。例えば、図11の例では、音韻「い」と「た」の区間に「A」のデータが記録されている。また、音韻「は」と「な」の区間に「S」のデータが記録されている。この表情トラックには、上記歌唱データ編集画面の各ノートバー毎に任意に指定された表情が反映される。

## 【0054】

ここで、歌唱データ編集画面においては、ノートバーがノートアタックの音の歌唱を指示するものであるかそれともノートランジッションの音の歌唱を指示するものであるかを区別することなく、「アクセント」、「ソフト（やわらかく）」、「レガート（なめらかに）」、「ポルタメント」のいずれの表情の指定であってもすることができた。しかしながら、レガート又はポルタメントをかけてノートアタックの音を歌唱することは現実にはありえない。従って、スコアデー

タ生成部 1 0 2 は、このような本来ありえない指定が行われているかを検出し、このような指定が行なわれているときはこれを無視する。

#### 【 0 0 5 5 】

図 5 のフローチャートにおいて、スコアデータ生成部 1 0 2 による歌唱スコアデータの生成 ( S 1 2 0 ) が終わると、スコアデータ更新部 1 0 3 は、歌唱スコアデータの状態トラックにデータを追加して、歌唱スコアデータを更新する ( S 1 3 0 ) 。この際、スコアデータ更新部 1 0 3 は、歌唱スコアデータの各音韻のコンテキストをコンテキスト判別部 1 0 4 に問い合わせる。そして、この判別結果に応じて、ノートアタックの音であることを示すコンテキストデータ或いはノートトランジッションの音であることを示すコンテキストデータを各音韻と関連付けて記録する。図 1 2 は、状態トラックにコンテキストデータが追加された歌唱スコアデータの一例を示す概念図である。同図では、音韻「さ」と「は」には、ノートアタックの音を示すコンテキストデータである「 a t t a c k 」が関連付けられ、音韻「い」と「た」と「な」には、ノートトランジッションの音を示すコンテキストデータである「 t r a n s i t i o n 」が関連付けられている。

#### 【 0 0 5 6 】

ここで、上記コンテキスト判別部 1 0 4 によるコンテキストの判別方法としては、歌唱スコアデータの音韻トラックに着目する第 1 の判別方法と、歌唱スコアデータのピッチトラックに着目する第 2 の判別方法とがある。

#### 【 0 0 5 7 】

まず、第 1 の判別方法の手順を以下に示す。まず、歌唱スコアデータの音韻トラックから、当該判別対象の音韻に時系列的に先行する直近の音韻の発音タイミングを特定する。そして、当該判断対象の音韻と上記先行する音韻との発音タイミングの差を求める。更に、当該判断対象の音韻と上記先行する音韻との差が予め設定された所定の間隔以内であれば、当該判断対象の音韻をノートトランジッションの音であると特定し、当該判断対象の音韻と上記先行する音韻との差が上記設定された所定の期間を超えると、或いは先行する音韻が存在していないときは、当該判断対象の音韻をノートアタックの音と特定する。

#### 【 0 0 5 8 】

第2の判別方法の手順を以下に示す。上述のように、歌唱スコアデータのピッチトラックには、各時刻において発音すべき各音韻の音声の基本周波数が記録される。したがって、まず、歌唱スコアデータのピッチトラックから、当該判別対象となる音韻のピッチの始点と、予め設定された所定期間を当該始点から時系列的にさかのぼった時刻とを特定する。そして、この特定した時刻から上記始点までのピッチトラックの区間に、ピッチを指定する値が記録されているか判断する。この区間にピッチを指定する値が記録されていれば当該判断対象の音韻をノートトランジションの音と特定し、記録されていなければノートアタックの音と特定する。

#### 【0059】

図5のフローチャートの説明に戻って、特徴パラメータ発生部105は、時刻  $t$  を進めながら、各時刻  $t$  における音韻に関する情報を歌唱スコアデータから取り出し、その音韻に対応した音声波形を合成するのに必要な特徴パラメータを *Timbre* データベース210から読み出してRAM120上に展開する(S140)。上述のように、*Timbre* データベース210は音韻名とピッチをインデックスとして編成されているので、歌唱スコアデータの音韻トラックの各音韻及びこれに対応するピッチトラックのピッチをキーとして、発音すべき音韻に対応した特徴パラメータを特定することが可能である。

#### 【0060】

特徴パラメータ発生部105は、歌唱スコアデータの時刻  $t$  における表情トラックの値に基づいて表情指定区間を特定し、表情テンプレートデータベース200から読み出した表情テンプレートデータをこの表情指定区間の特徴パラメータ及びピッチに適用する(S150)。このステップ150の表情テンプレートデータ適用処理を、図13のフローチャートを参照して更に詳細に説明する。

#### 【0061】

ステップ151において特徴パラメータ発生部105は、時刻  $t$  において何らかの表情が表情トラックで指定されているか判断する。時刻  $t$  において「A」、「S」、「R」、「P」のいずれかのデータが表情トラックに記録されているときは、表情の指定があると判断する。表情が指定されているときは、ステップ1

52に進み、表情が指定されていないときは、ステップ151に再び戻り、時刻  $t$  を進めて、以降の処理を実行する。

#### 【0062】

ステップ152において、特徴パラメータ発生部105は、表情トラックにおいて時刻  $t$  における表情と同じ表情属性を持つ領域の開始時刻及び終了時刻（例えば時刻  $t$  における表情属性がアクセントを示す「A」である場合には、その「A」が始まる時刻と終了する時刻）を求める。この開始時刻と終了時刻の間が表情テンプレートデータを適用する表情指定区間になる。

#### 【0063】

ステップ153において、特徴パラメータ発生部105は、状態トラックの時刻  $t$  におけるデータが「attack」のコンテキストデータであるかそれとも「transition」のコンテキストデータであるかを判断する。「attack」のコンテキストデータが記録されているときはステップ154に進み、「transition」のコンテキストデータが記録されているときはステップ155に進む。

#### 【0064】

ステップ154では、特徴パラメータ発生部105が、ノートアタック表情テンプレートデータベース220から表情テンプレートデータを読み出す。上述のように、ノートアタック表情テンプレートデータベース220内には、アクセントテンプレートデータベース221とソフトテンプレートデータベース222とを備えており、これらデータベースはいずれも、音韻名と代表ピッチをインデックスとして編成されている。従って、ステップ154では、まず、時刻  $t$  における表情トラックの表情属性に対応するデータベース（例えば表情属性が「A」ならアクセントテンプレートデータベース221）を特定し、時刻  $t$  における音韻トラック及びピッチトラックの値に対応するテンプレートデータをこのデータベースから更に特定する。

#### 【0065】

一方、ステップ155では、特徴パラメータ発生部105が、ノートランジション表情テンプレートデータベース230から表情テンプレートデータを読

み出す。上述のように、ノートランジション表情テンプレートデータベース 230 内には、アクセントテンプレートデータベース 231 とソフトテンプレートデータベース 232 とレガートテンプレートデータベース 233 とポルタメントテンプレートデータベース 234 とを備えられており、これらデータベースはいずれも、先頭音韻名、後続音韻名、代表ピッチをインデックスとして編成されている。従って、ステップ 155 では、時刻  $t$  における表情トラックの値に対応するデータベース（例えば「A」ならアクセントテンプレートデータベース 231）を特定し、音韻トラックに記憶された時刻  $t$  における音韻（すなわち、図 4 における後続音韻）とその直前の音韻（すなわち、図 4 における先頭音韻）及びピッチトラックに記録された時刻  $t$  におけるピッチ（すなわち、図 4 における代表ピッチ）をインデックスとするテンプレートデータをこのデータベースから更に特定する。

#### 【0066】

ステップ 156 において、特徴パラメータ発生部 105 は、ステップ 154 または 155 において読み出した表情テンプレートデータを上記表情指定区間に対応する時間長に伸張し、当該表情指定区間のピッチ及び特徴パラメータと当該伸張した表情テンプレートデータの値とを交換する。

#### 【0067】

以上の処理を時刻  $t$  を増加させながら繰り返し行うことにより、アクセント、レガートといった表情の指定に応じた特徴パラメータ及びピッチが、演奏時間の経過に従って生成される。

#### 【0068】

以上説明した図 13 の処理が終了すると、図 5 のフローチャートにおいて、特徴パラメータ発生部 105 は、音韻テンプレートデータベース 240 から読み出した音韻テンプレートデータを特徴パラメータ及びピッチに適用する（S160）。この音韻テンプレートデータの適用は、歌唱スコアデータの音韻トラックの値から音韻の遷移区間を特定し、音韻テンプレートデータベース 240 から読み出した音韻テンプレートデータをこの遷移区間に対応する時間長に伸張し、上記遷移区間のピッチ及び特徴パラメータに当該伸張した音韻テンプレートデータの

値を加算するといった処理によって実現する。但し、このような適用手順は従来技術の範疇であるため詳細な説明は省略する。

#### 【0069】

特徴パラメータ発生部105は、状態テンプレートデータベース250から読み出した状態テンプレートデータを特徴パラメータ及びピッチに適用する（S170）。この状態テンプレートデータの適用は、歌唱スコアデータの状態トラック及びピッチトラックの値から音韻の立ち上がり或いは遷移区間を特定し、状態テンプレートデータベース250から読み出した状態テンプレートデータをこれら区間に対応する時間長に伸張し、上記各区間のピッチ及び特徴パラメータに当該伸張した状態テンプレートデータの値を加算するといった処理によって実現する。但し、このような適用手順は従来技術の範疇であるため詳細な説明は省略する。

#### 【0070】

最後に、歌唱合成部106は、以上のようにして最終的に得られた特徴パラメータ及びピッチに基づいて、デジタル音声データを合成する（S180）。そして、この合成された音声データは、DAC170でアナログデータに変換され、サウンドシステム180から発音されることになる。

#### 【0071】

以上説明したように、本実施形態によれば、歌唱データの合成のためのデータ入力を行うユーザは、所望の区間に表情付けを行いたい場合、その区間が置かれたコンテキストを意識することなく、その区間に付けるべき表情を指定するだけで、その区間が置かれているコンテキストおよびユーザが指定した表情に合った適切な歌唱データが合成される。

#### 【0072】

### B 第2の実施の形態

#### <実施形態の構成>

本実施形態に係る歌唱合成装置の物理的構成は、上記第1の実施の形態と同様であり、図を参照しての再度の説明を省略する。

#### 【0073】



図14は、当該歌唱合成装置の論理的構成を示すブロック図である。同図左側には、CPU100がその役割を担う各部の構成を示し、同図右側には、記憶装置200に編成されたデータベースの構成を示す。

#### 【0074】

まず、記憶装置200にインストールされた歌唱合成プログラムを実行することによって、CPU100は、インターフェース制御部101、スコアデータ生成部102、コンテキスト判別部104、特徴パラメータ発生部105、及び歌唱合成部106としての役割を担う。上記第1の実施の形態と異なり、スコアデータ更新部103は備えていない。

#### 【0075】

インターフェース制御部101の機能は上記第1の実施の形態と同様であり、図6に示した歌唱データ編集画面をディスプレイ140に表示する。また、スコアデータ生成部102の機能も上記第1の実施の形態と同様である。

#### 【0076】

本実施形態におけるコンテキスト判別部104は、特徴パラメータ発生部105からの問い合わせを受け、歌唱スコアデータに記録された特定の音韻のコンテキストを判別する。特徴パラメータ発生部105は、特徴パラメータをデータベースから読み出すと共に、上記コンテキスト判別部104の判別結果に応じたテンプレートデータを読み出して当該特徴パラメータに適用する。

#### 【0077】

歌唱合成部106の機能は上記第1の実施の形態と同様である。

#### 【0078】

また、データベースの編成も上記第1の実施の形態と同様である。

#### 【0079】

#### ＜実施形態の動作＞

以下、上記構成を有する歌唱合成装置の動作について説明する。図15は、本実施形態に係る歌唱合成装置の動作の概要を示すフローチャートである。

#### 【0080】

CPU100は、GUIを介して歌唱合成の実行指示を受け取ると、記憶装置

200 から歌唱合成プログラムを読み出して実行する。この歌唱合成プログラムの実行過程では、図15に示す処理が行われる。この図15において、ステップS210～S220およびステップS240～S270の処理内容は、上述した第1の実施形態の図5におけるステップS110～S120およびステップS150～S180の処理内容と同じである。図5における処理では、ステップS130において、歌唱スコアデータに状態トラックのデータを追加する更新処理が行われた。これに対し、図15に示す処理は、このステップS130に相当する処理を含んでいない。その代わりに、図15においてステップS230において実行される処理は、図13に示すものから図16に示すものに置き換えられている。本実施形態と上記第1の実施形態との相違は以上の点にある。

#### 【0081】

図16において、ステップS241～S242およびステップS244～S246の処理内容は、図13におけるステップS151～S152およびステップS154～S156の処理内容と同じである。図16では、図13におけるステップS153がステップS243aおよびS243bに置き換えられている。従って、説明の重複を避けるため、以下、これらのステップのみについて説明する。

#### 【0082】

まず、ステップ243aにおいて、特徴パラメータ発生部105は、時刻tを終期とする一定時間内に属するデータを歌唱スコアデータの音韻トラックとピッチトラックから取り出し、コンテキスト判別部104に引き渡して、時刻tにおける音韻のコンテキストを問い合わせる。

#### 【0083】

次にステップ243bにおいて、コンテキスト判別部104は、特徴パラメータ発生部105から引き渡されたデータに基づいて、時刻tにおける音韻のコンテキストを判別する。当該コンテキスト判別部104によって当該音韻がノートアタックの音であると判別されたときは、ステップ244に進み、ノートランジションの音であると判別されたときは、ステップ245に進む。

#### 【0084】

以上説明した第2の実施の形態は、歌唱スコアデータに記録された各音韻のコンテキストを判別するタイミングが上記第1の実施の形態と異なる。上記第1の実施の形態では、特徴パラメータの生成を開始する前の段階で各音韻のコンテキストを判別し、この判別結果に応じて「attack」または「Transition」のいずれかのコンテキストデータを歌唱スコアデータに記録するようになっていた。これに対し、本実施形態では、各音韻のコンテキストを特定するデータを何ら有しない歌唱スコアデータを特徴パラメータ発生部105が取得する。そして、特徴パラメータ発生部105がテンプレートデータをデータベースから読み出す段階で、各音韻のコンテキストの判別が行われるようになっている。このような構成をとる本実施形態では、状態トラックを歌唱スコアデータに設ける必要がなくなるため、歌唱スコアデータの容量を小さくすることができる。

#### 【0085】

#### C 変形例

以上この発明の一実施形態について説明したが、上記実施形態はあくまでも例示であり、上記実施形態に対しては、様々な変形を加えることができる。変形例としては、例えば以下のようなものが考えられる。

#### 【0086】

##### < C-1 変形例1 >

上記実施の形態における歌唱データ編集画面では、「アクセント」、「ソフト（やわらかく）」、「レガート（なめらかに）」、「ポルタメント」のいずれかの表情を各ノートバー毎に指定するが、この指定は、ノートアタックの音の歌唱を指示するノートバーであるとノートトランジションの音の歌唱を指示するノートバーであるとを問わずに行うことができるようになっていた。そして、ノートアタックの音にレガートの表情を指定するような本来ありえない指定が行われているかを、スコアデータ生成の際、或いは特徴パラメータ生成の際に判断し、このような指定を無視するような構成にしていた。

#### 【0087】

これに対し、図17に示すような論理的構成をとることによって、本来ありえない指定を行えないように、インターフェース制御部101が上記歌唱データ編

集画面からの入力を制限するようにしてもよい。この入力の制限の方法は以下のようなものが想定できる。まず、上記歌唱データ編集画面に形成されたノートバーについての表情の指定が入力されると、インターフェース制御部101が、当該ノートバーがノートアタックの音の歌唱を指示するものかそれともノートランジッションの音の歌唱を指示するものかをコンテキスト判別部104に問い合わせる。そして、この判別の結果、当該ノートバーがノートアタックの音の歌唱を指示するものであったとき、当該インターフェース制御部101は、「この音符は立ち上がり音なので、レガート及びポルタメントをかけることができません。」といったメッセージを表示する。

#### 【0088】

##### < C-2 変形例2 >

上記実施の形態における歌唱スコアデータは、音韻トラック、ピッチトラック、表情トラックの3トラック、或いはこれに状態トラックを加えた4トラックで構成したが、別のトラック構成をとってもよい。例えば、音声の強さを示すパラメータである各時刻のダイナミクス値を記録するトラックや、唇の開き具合を示すパラメータである各時刻のオープニング値を記録するトラックを更に加えることによって、より一層人間の肉声に近い歌唱音声を再現することが可能である。

#### 【0089】

##### 【発明の効果】

以上説明したように、本発明によると、音声情報に含まれる各音韻が立ち上がり音であるかそれとも非立ち上がり音であるかを判別する判別手段を備え、立ち上がり音に適用するテンプレートデータと、非立ち上がり音に適用するテンプレートデータとを個別に用意しておく。そして、音声情報が入力されたときは、この音声情報に適用するテンプレートデータを上記判別手段の判別結果に応じて自動的に特定する。従って、ユーザは、各音韻が立ち上がり音であるか非立ち上がり音であるかを意識することなく、表情を持たせた音声を合成するための音声情報を容易に生成できる。

##### 【図面の簡単な説明】

【図1】 歌唱合成装置の物理的構成を示すブロック図である。

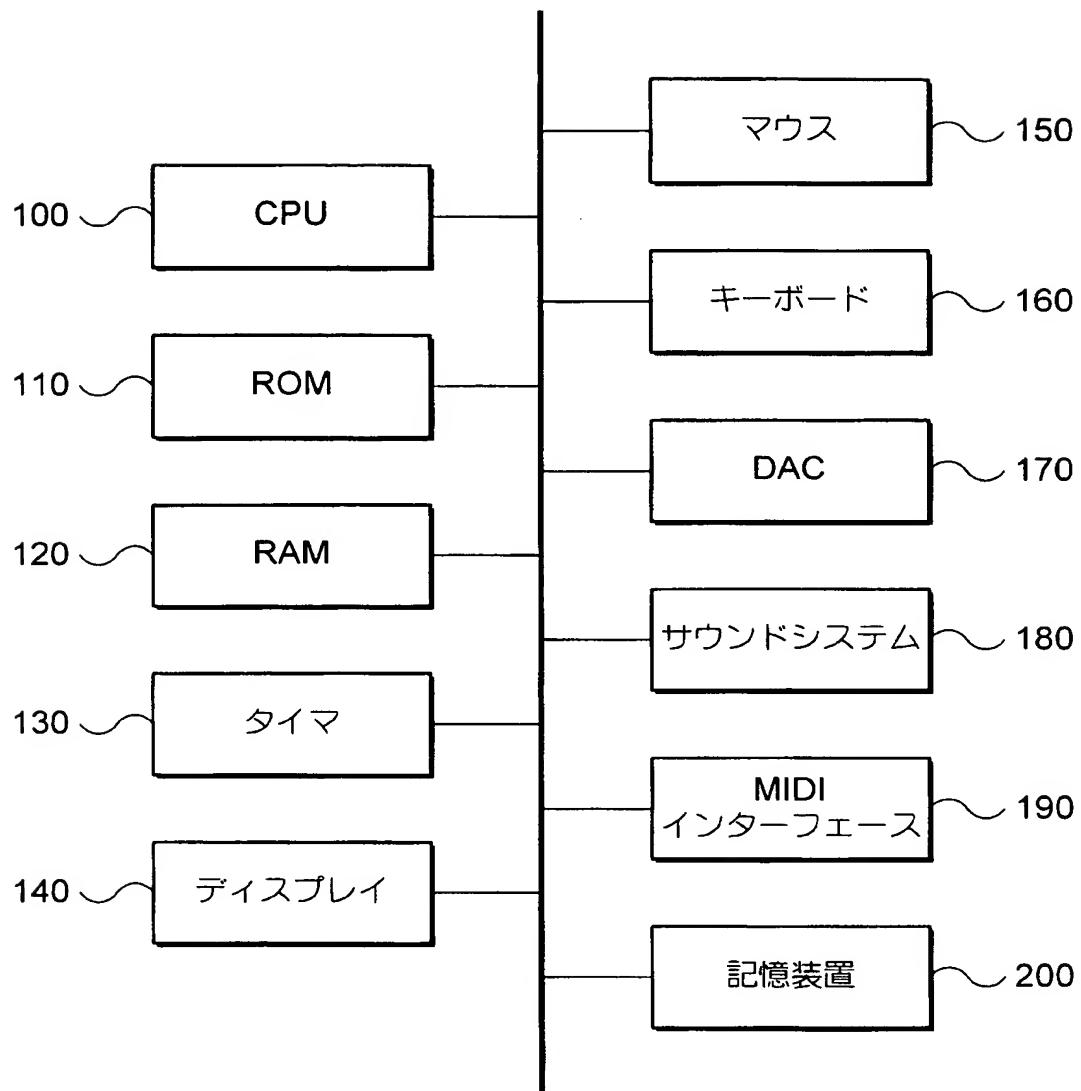
- 【図 2】 歌唱合成装置の論理的構成を示すブロック図である。
- 【図 3】 テンプレートデータベースのデータ構造の一例である。
- 【図 4】 テンプレートデータベースのデータ構造の一例である。
- 【図 5】 実施形態の動作を示すフローチャートである。
- 【図 6】 歌唱データ編集画面の一例である。
- 【図 7】 歌詞入力欄の一例である。
- 【図 8】 表情選択リストの一例である。
- 【図 9】 ノートバーの入力例である。
- 【図 10】 歌詞の入力例である。
- 【図 11】 歌唱スコアデータの一例である。
- 【図 12】 歌唱スコアデータの一例である。
- 【図 13】 表情テンプレート適用処理を示すフローチャートである。
- 【図 14】 歌唱合成装置の論理的構成を示すブロック図である。
- 【図 15】 実施形態の動作を示すフローチャートである。
- 【図 16】 表情テンプレート適用処理を示すフローチャートである。
- 【図 17】 歌唱合成装置の論理的構成を示すブロック図である。

【符号の説明】

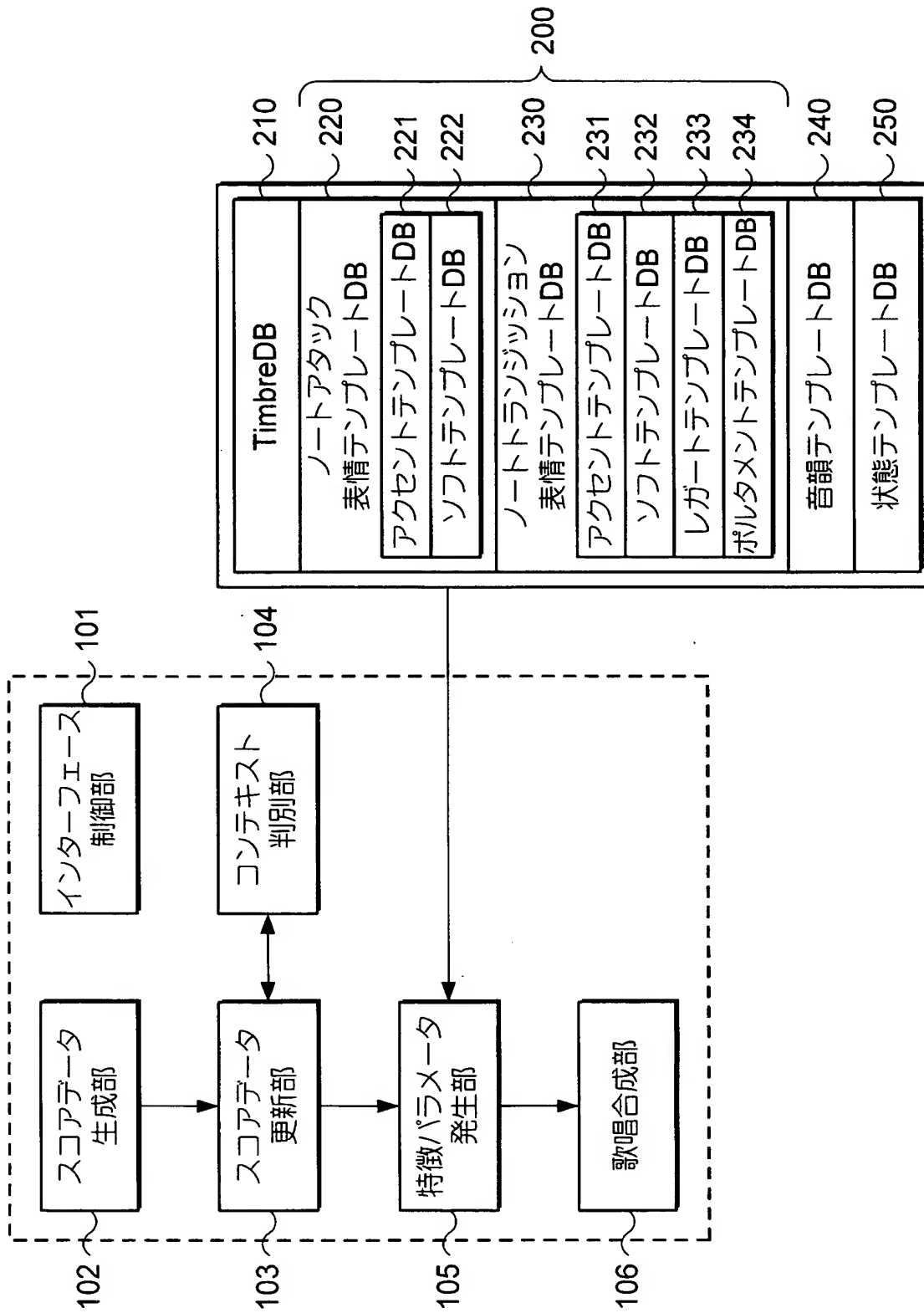
100…CPU、110…ROM、120…RAM、130…タイマ、140…ディスプレイ、150…マウス、160…キーボード、170…DAC、180…サウンドシステム、190…MIDI インターフェース、200…記憶装置、101…インターフェース制御部、102…スコアデータ生成部、103…スコアデータ更新部、104…コンテキスト判別部、105…特徴パラメータ発生部、106…歌唱合成部、210…Timbre データベース、220…ノートアタック表情テンプレートデータベース、230…ノートトランジッション表情テンプレートデータベース、240…音韻テンプレートデータベース、250…状態テンプレートデータベース。

【書類名】 図面

【図 1】



【図 2】



【図 3】

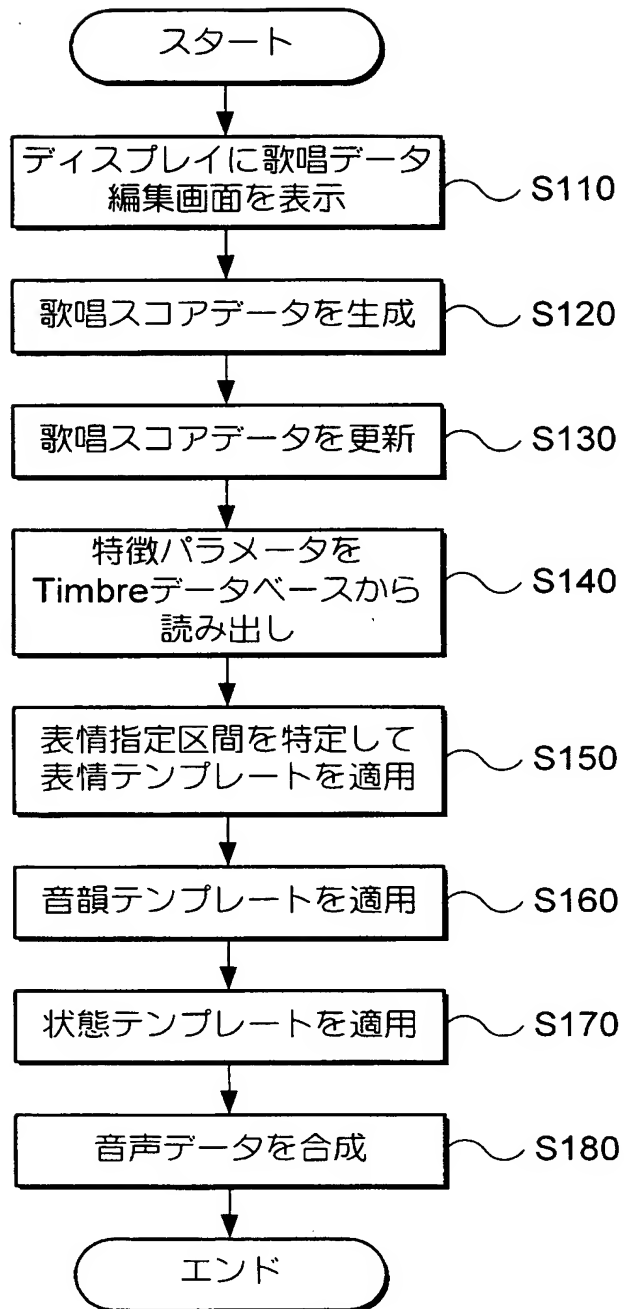
音韻名	代表ピッチ	テンプレート
/あ/	200	{P(t),Pitch(t),T}
/あ/	300	{P(t),Pitch(t),T}
/あ/	400	{P(t),Pitch(t),T}
/い/	200	{P(t),Pitch(t),T}
/い/	300	{P(t),Pitch(t),T}
⋮	⋮	⋮

【図 4】

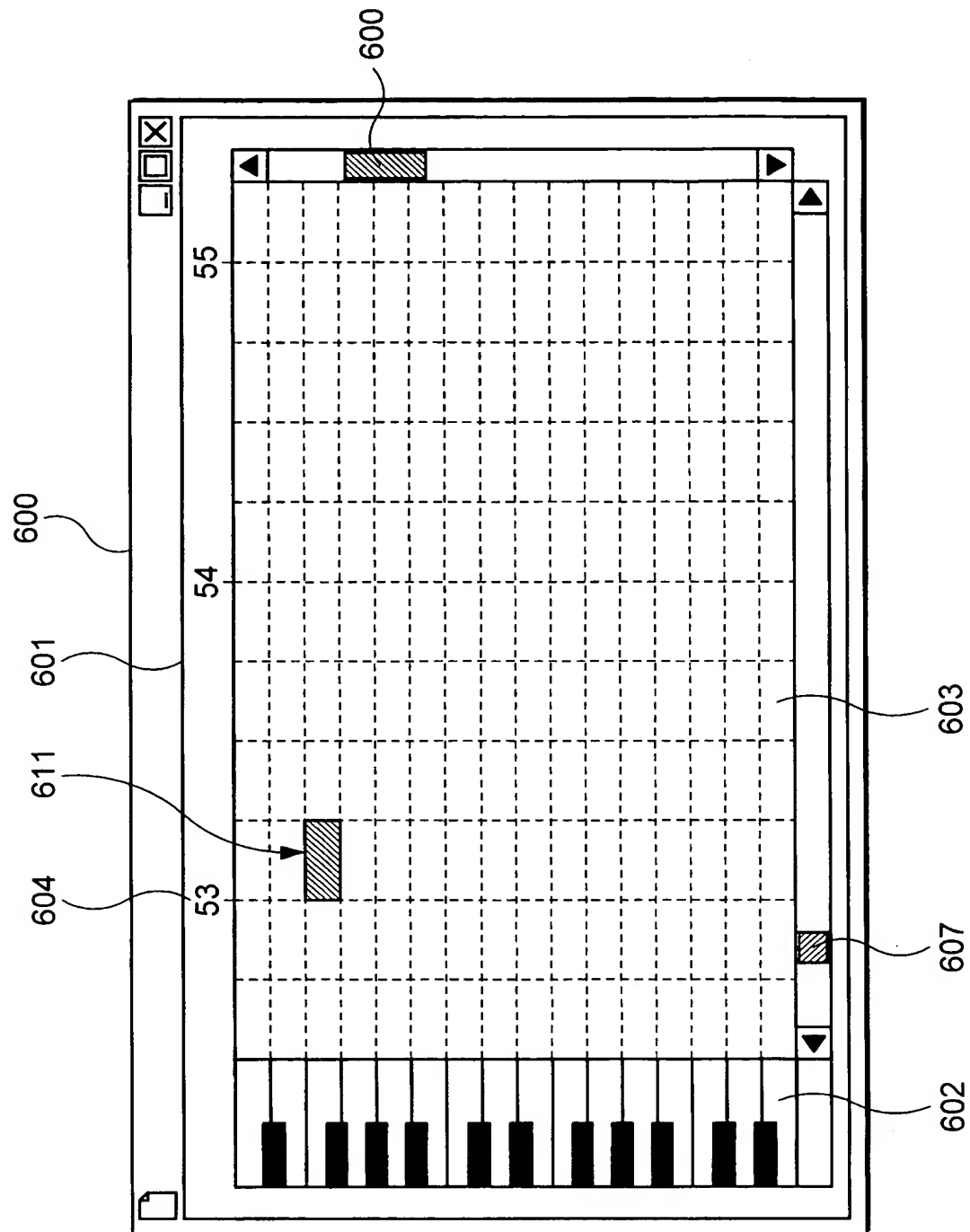
先頭音韻名	後続音韻名	代表ピッチ	テンプレート
/あ/	/あ/	200	{P(t),Pitch(t),T}
/あ/	/あ/	300	{P(t),Pitch(t),T}
/あ/	/あ/	400	{P(t),Pitch(t),T}
/あ/	/い/	200	{P(t),Pitch(t),T}
/あ/	/い/	300	{P(t),Pitch(t),T}
⋮	⋮	⋮	⋮



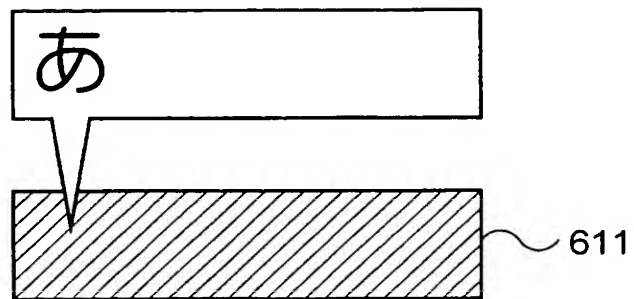
【図 5】



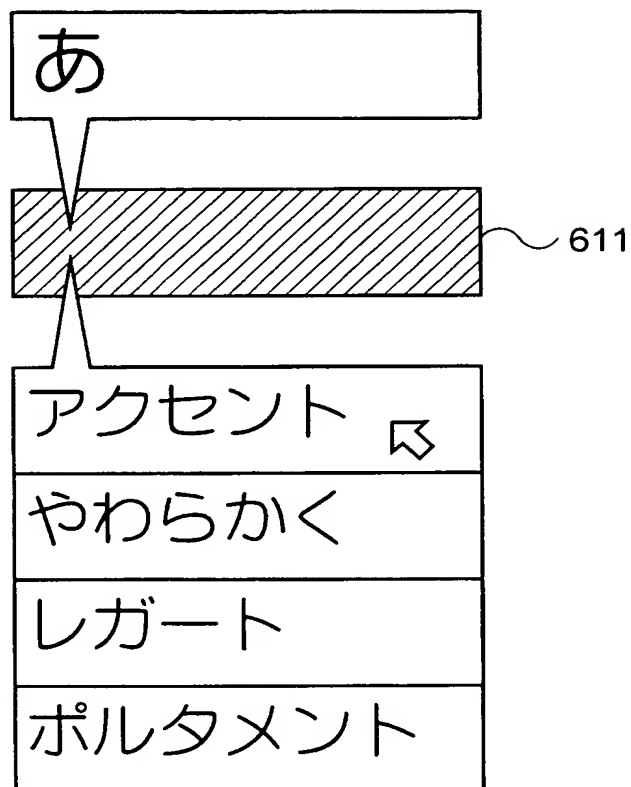
【図 6】



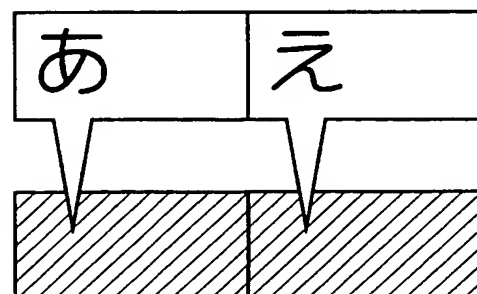
【図 7】



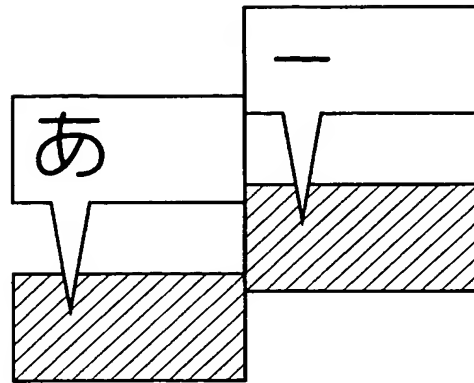
【図 8】



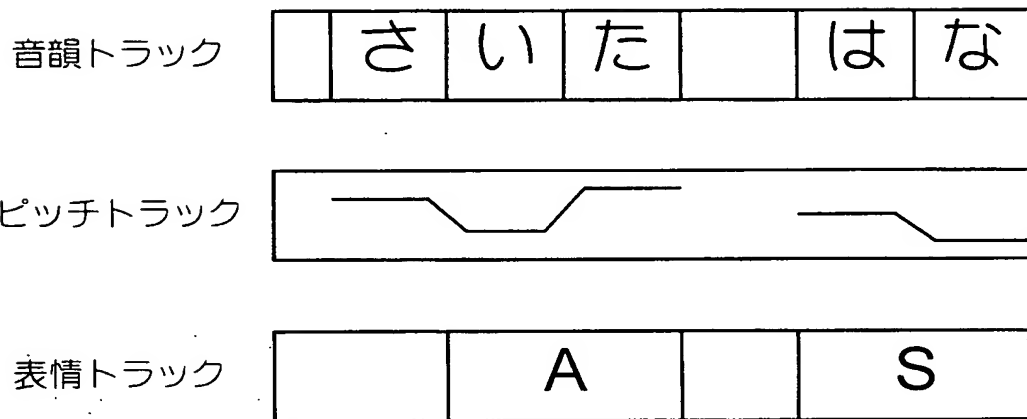
【図 9】



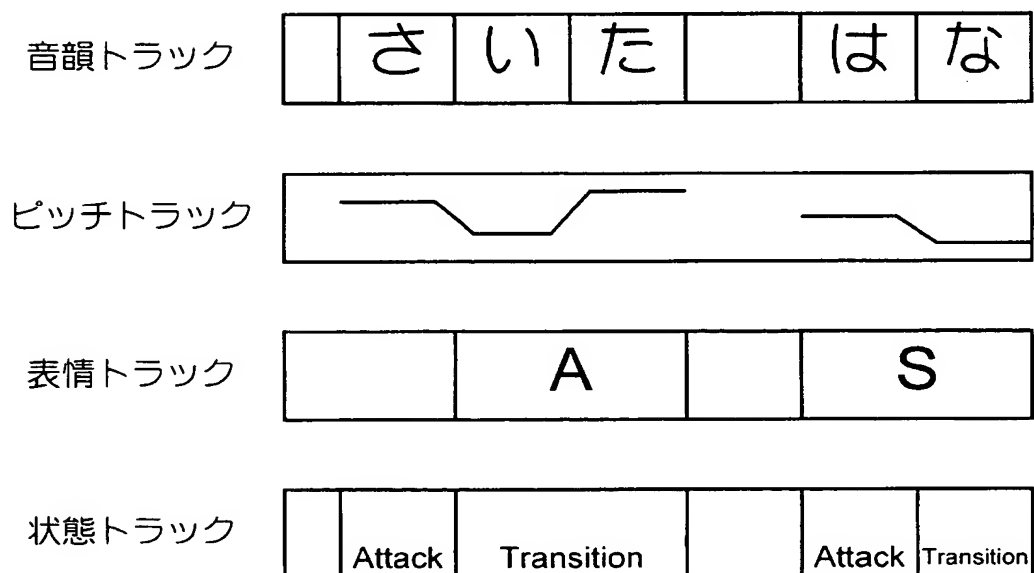
【図 10】



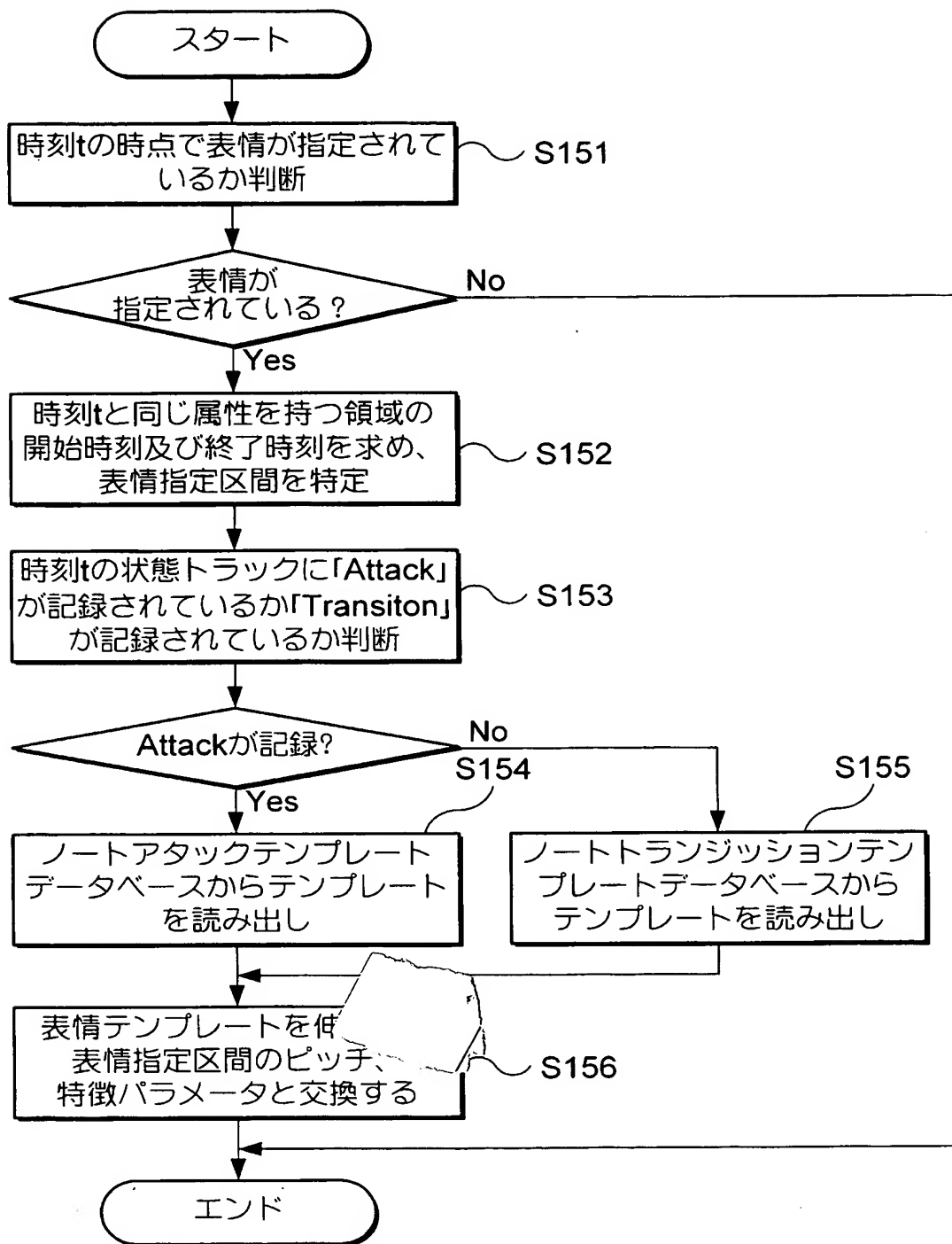
【図 11】



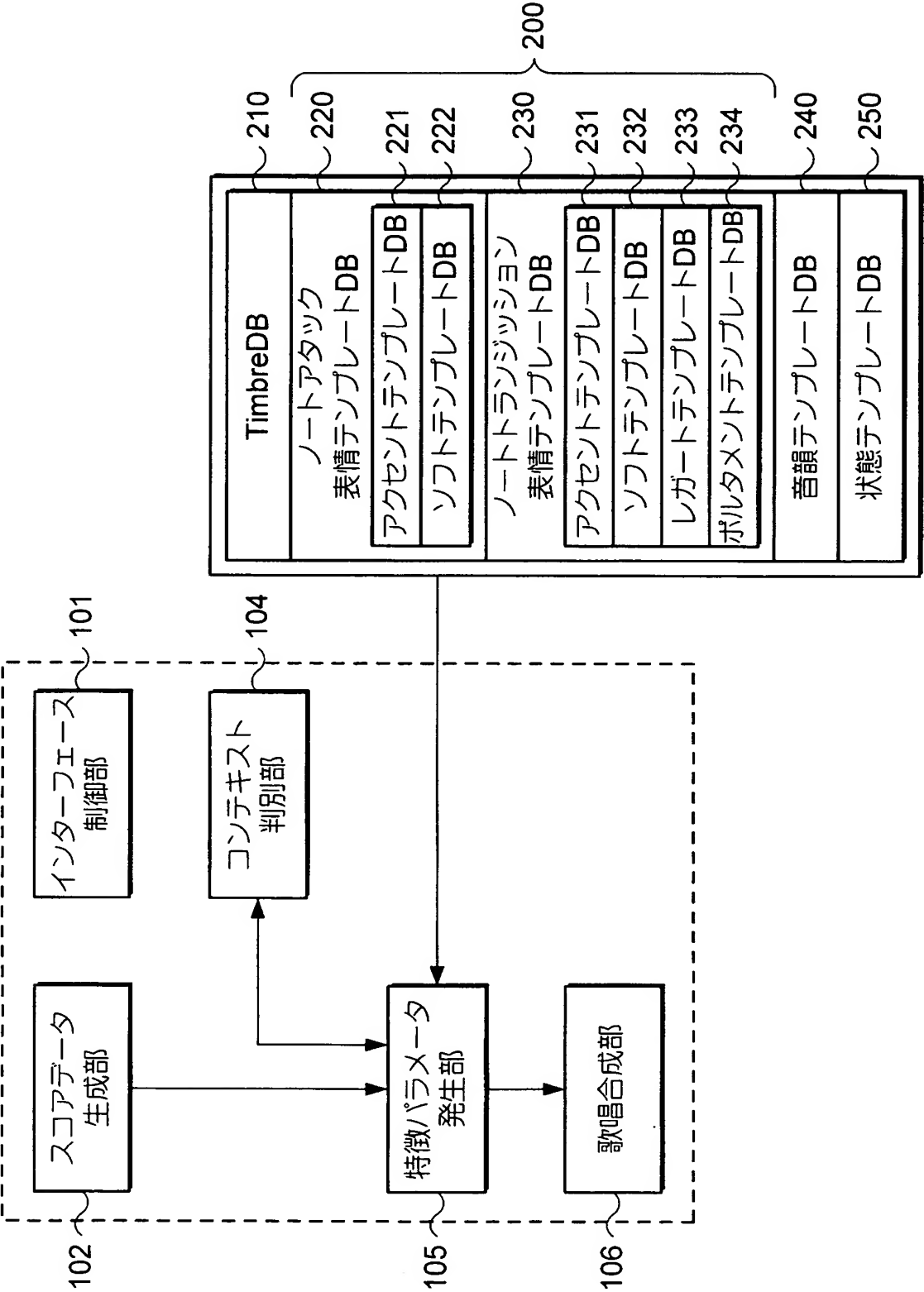
【図 12】



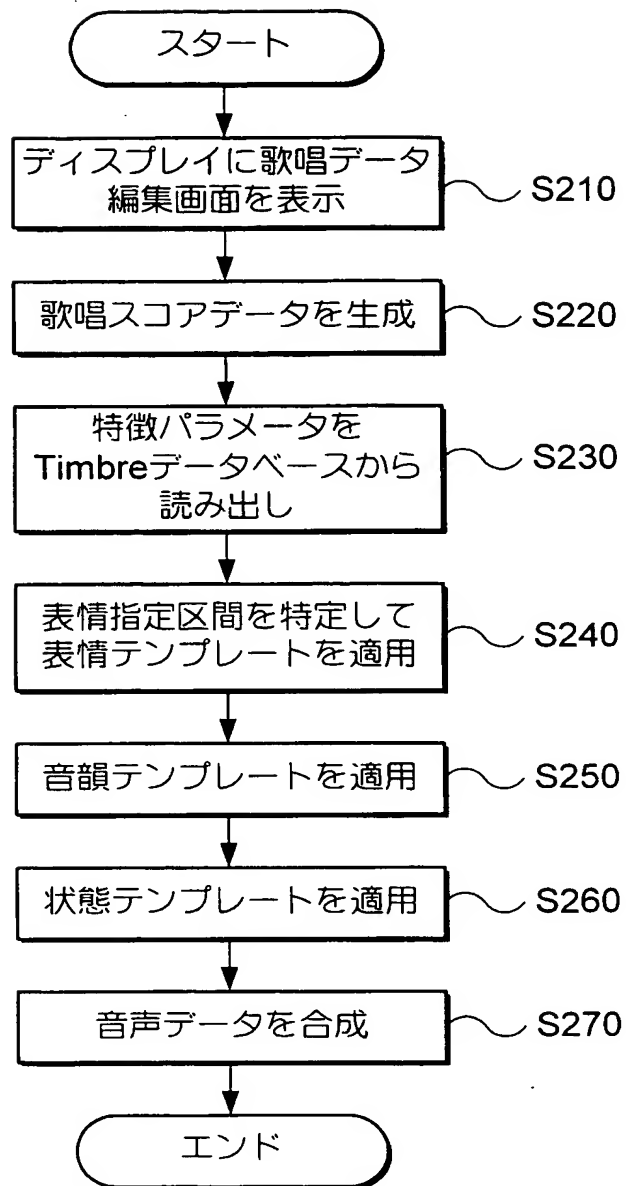
【図13】



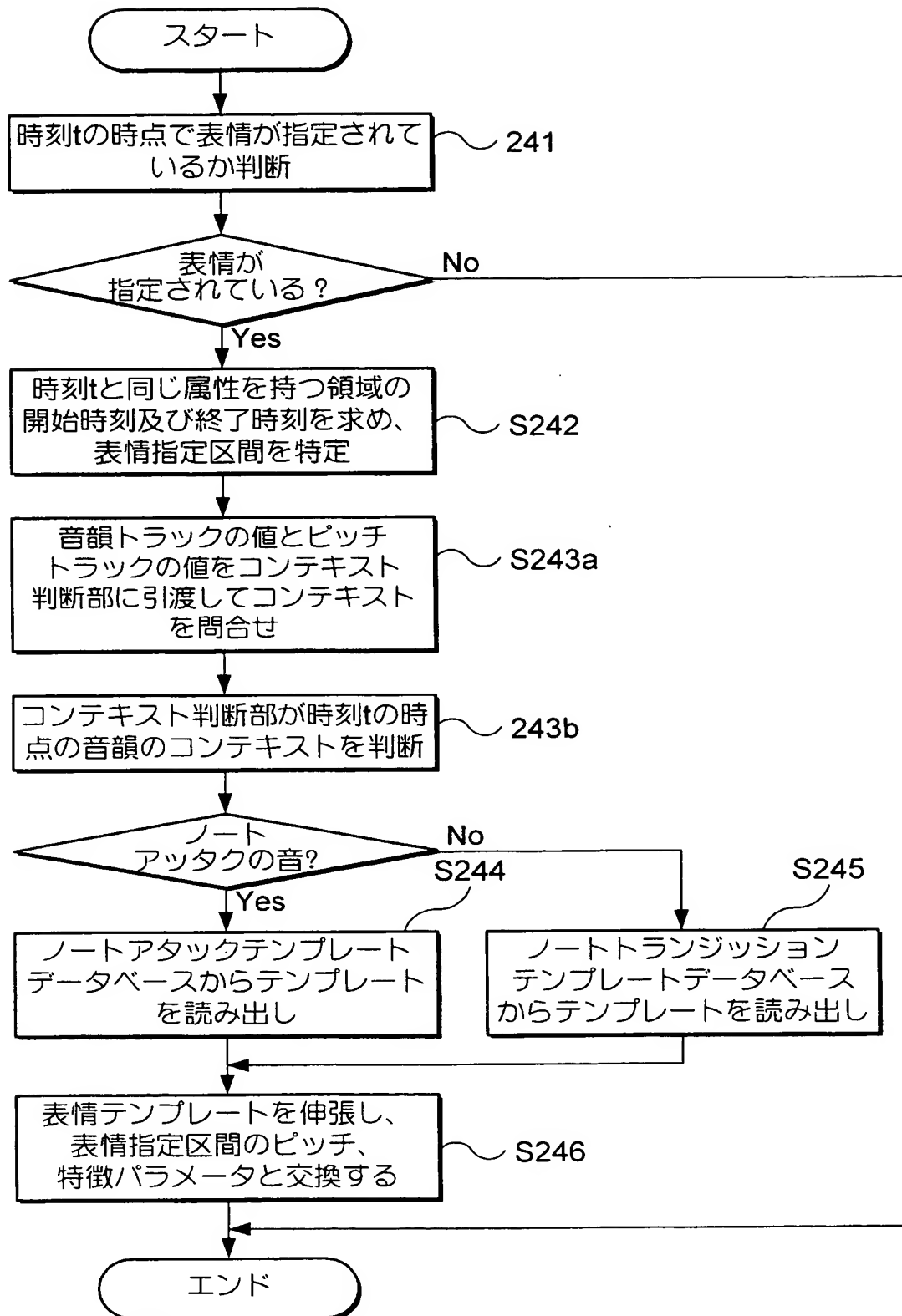
【図 14】



【図 15】

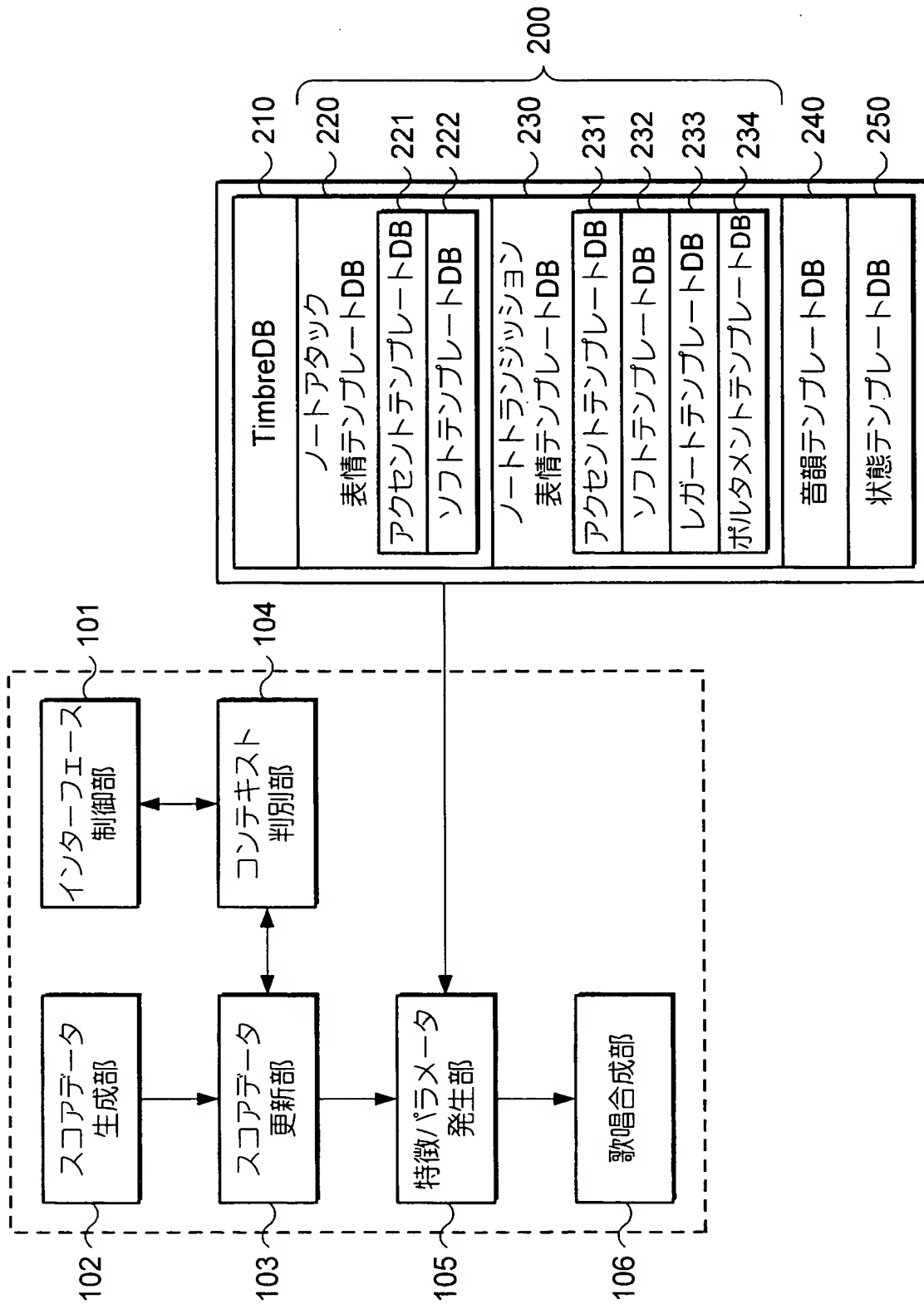


【図 16】





【図 17】



【書類名】 要約書

【要約】

【課題】 歌唱データを作成するユーザが、表情を与える区間の前に連続する音符があるかといったことを意識しなくても、適切な表情テンプレートが適用されるような歌唱合成装置を提供すること。

【解決手段】 立ち上がり音の特徴量の時間的変化をテンプレートデータとして記憶するノートアタック表情テンプレートデータベース 2 2 0 と、非立ち上がり音の特徴量の時間的変化をテンプレートデータとして記憶するノートトランジション表情テンプレートデータベース 2 3 0 とを備える。更に、コンテキスト判別手段 1 0 4 を備え、この手段による音韻の判別結果に応じて、上記いずれのデータベースから読み出したテンプレートを用いて音声を合成するかを決定する。

【選択図】 図 2

特願 2 0 0 3 - 0 5 5 8 9 8

出 願 人 履 歴 情 報

識別番号

[ 0 0 0 0 0 4 0 7 5 ]

1 . 変更年月日

1 9 9 0 年 8 月 2 2 日

[変更理由]

新規登録

住 所

静岡県浜松市中沢町 1 0 番 1 号

氏 名

ヤマハ株式会社